Folding Kinetics for the Conformational Switch between Alternative RNA Structures

Song Cao,[†] Boris Fürtig,[‡] Harald Schwalbe,[‡] and Shi-Jie Chen*,[†]

Department of Physics and Astronomy and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, and Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University, Max-von-Laue-Strasse 7, D-60438 Frankfurt/Main, 44780, Germany

Received: August 20, 2010

Transitions between different conformational states, so-called conformational switching, are intrinsic to RNA catalytic and regulatory functions. Often, conformational switching occurs on time scales of several seconds. In combination with the recent real-time NMR experiments (Wenter et al. *Angew. Chem. Int. Ed.* **2005**, *44*, 2600; Wenter et al. *ChemBioChem* **2006**, *7*, 417) for the transitions between bistable RNA conformations, we combine the master equation method with the kinetic cluster method to investigate the detailed kinetic mechanism and the factors that govern the folding kinetics. We propose that heat capacity change (ΔC_p) upon RNA folding may be important for RNA folding kinetics. In addition, we find that, for tetraloop hairpins, noncanonical (tertiary) intraloop interactions are important to determine the folding kinetics. Furthermore, through theory−experiment comparisons, we find that the different rate models for the fundamental steps (i.e., formation/disruption of a base pair or stack) can cause contrasting results in the theoretical predictions.

I. Introduction

RNA folding often involves misfolded states and alternative conformations.[1–3] Due to the strong stabilizing force from base pairing and base stacking, detrapping from a misfolded intermediate (time scale ≥ seconds) is usually slow compared to the folding rate of an RNA hairpin (time ∼ microseconds) from an unstructured state.[4–8] For example, the time scale for the conformational switch between the two alternative folds for the splicing leader RNA from *Leptomonas collosoma* takes about a second.[9] Folding of group I or II intron undergoes multiple intermediates and can take several minutes.[10–12]

In order to understand the detailed kinetic mechanisms of folding, several kinetic models have been developed in recent years,[6,13–15] including simulational methods[13–18] and the master equation (analytical) method.[6,19,20] The simulational method applied to RNA hairpins can be useful for elucidating the detailed folding pathways.[16,17] For example, the recent simulations on a simple pseudoknot structure show that the folding kinetics is correlated to the thermodynamic stability of the helix stems in the pseudoknot.[18] Monte Carlo simulation[13–15] has also been used to predict the slow folding kinetics for hepatitis delta virus (HDV) ribozyme,[13] a more complex pseudoknot structure. The computational results provided useful relationships between the kinetic folding pathways and the HDV activity.[13] However, the Monte Carlo simulations can potentially have the problem of incomplete sampling of the conformational ensemble. In contrast, the master equation method, which allows for a deterministic solution for the kinetics, is based on the complete conformational ensemble. The analytical solution to the master equations can give stable predictions for the long-time kinetics.[6,19] This is of particular importance, since RNA as a chain molecule has a large ensemble of conformations. The large size of the system would prohibit the analytical solution to the equations. To solve the master equation for large systems, several methods have been developed, such as the exact stochastic simulation method as a simulational approach to the solution of the equation.[14,15] Another way to effectively reduce the conformational ensemble is to classify the preequilibrated conformational subsets into clusters and to treat the original conformational ensemble as a collection of clusters.

In the present study, with a combination of the master equation and kinetic cluster theory, we investigate the slow folding kinetics for the conformational switch between alternative RNA structures. Specifically, we use two experimentally studied short RNAs[21,22] as the model system. We aim to explore the detailed folding pathways and transition states, and the factors that govern RNA folding kinetics. Both RNAs have a tetraloop hairpin as a kinetic intermediate. We find that the heat capacity change ΔC_p and the stabilizing energetics (2−3 kcal/mol)[23] of the tetraloop can significantly influence the rates of the conformational switch. Using the kinetic cluster method,[24,25] we can unveil the detailed folding pathways and the rate-limiting step. For example, we find that the slow folding rate between two hairpin structures is mainly derived from (a) the large enthalpic cost for breaking the base stack that closes to the tetraloop and (b) the small forward reaction probabilities for the states on the pathway.

II. Model and Theory

Master Equation Method. The master equation describes how the population of the conformations evolves with time; i.e., the kinetics for the fractional population (or the probability) $p_i(t)$ for the ith state ($i = 0, ..., \Omega - 1$, where Ω is the total number of chain conformations) is determined by the following master equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}p_i(t) = \sum_{j=0}^{\Omega-1} (k_{j\to i}p_j(t) - k_{i\to j}p_i(t)) \qquad (1)$$

where $k_{j\to i}$ and $k_{i\to j}$ are the rate constants for the respective transitions. Equation 1 is equivalent to the following expression:

* Corresponding author. Fax: 573-882-4195. E-mail: chenshi@missouri.edu.
† University of Missouri.
‡ Johann Wolfgang Goethe-University.

10.1021/jp107912s
Published on Web 10/01/2010

$d\mathbf{p}(t)/dt = \mathbf{M}\cdot\mathbf{p}(t)$, where M is the rate matrix, defined as $M_{ij} = k_{i\to j}$ for $i \neq j$ and $M_{ij} = -\sum_{l\neq i} k_{il}$ for $i = j$, and $\mathbf{p}(t)$ is the fractional populational vector col $[p_0(t), p_1(t), ..., p_{\Omega-1}(t)]$.

By diagonalizing the rate matrix M, we can obtain the eigenvalues λ_m and the eigenvectors $\mathbf{n}_m$ ($m = 0, 1, ..., \Omega - 1$). The populational kinetics $\mathbf{p}(t)$ is written as a sum over all the possible eigenmodes of the rate matrix:

$$\mathbf{p}(t) = \sum_{m=0}^{\Omega-1} C_m \mathbf{n}_m e^{-\lambda_m t} \quad (2)$$

where C_m is the coefficient that is dependent on the initial condition. The eigenvalue spectrum contains a static equilibrium state corresponding to the eigenvalue $\lambda_0 = 0$. We denote the first and second nonzero eigenvalues as λ_1 and λ_2, respectively. A large gap between λ_1 and λ_2 indicates a single exponential kinetics with the rate constant equal to λ_1. In the rate model, we allow only one base pair or one base stack to be formed or broken in a kinetic transition. The transition rate for opening or closing a base pair is given by

$$k_{\text{open}} = k_0^- e^{\Delta H/k_B T}, \quad k_{\text{close}} = k_0^+ e^{\Delta S/k_B} \quad (3)$$

where k_B is the Boltzmann constant and T is the temperature. We assume that the rate constants k_0^- and k_0^+ for opening and closing a (single) base pair are the same (rate constant k_0). In the calculation, k_0 is a prefactor fitted from the experimental data.[7] ΔH and ΔS are the enthalpy and entropy changes for the formation of the base stack.

The above (ΔH, ΔS) rate model is based on the different assumptions about the transition states for the formation and disruption of a base-pair/stack. For the formation of a base-pair/stack, we assume that the transition state occurs when the participating nucleotides have been restricted to the base-paired positions but not yet stabilized by the hydrogen bonding/base stacking interactions. However, for the disruption of a base-pair/stack, the transition state occurs when the hydrogen bonding/base stacking interactions have been disrupted but the nucleotides are not yet liberated from the base-paired positions. Such a transition state in the rate model is in accordance with the results from the molecular dynamics simulations.[26]

Kinetic Cluster Method. The large number of chain conformations (Ω) causes a large size ($\Omega \times \Omega$) of the rate matrix for the master equation. However, some conformations (microstates) may quickly interconvert and preequilibrate; these conformations can be treated as a macrostate (kinetic cluster). With the original complete conformational ensemble divided into kinetic clusters, the overall kinetics of the system would be determined by the transitions between the different clusters (kinetic cluster theory).[24,25,27] For example, if formation/disruption of a particular base stack denoted as s^* is a single rate-limiting step for the folding, the original conformational ensemble can be reduced into two clusters U and N, each consisting of all the conformations with or without s^*, respectively.

Transition between two clusters ($U \rightleftharpoons N$) is realized by transitions $U_i \rightleftharpoons N_i$ (called micropathways) between conformations (microstates) U_i in cluster U and N_i in cluster N. Here, $U_i \rightleftharpoons N_i$ corresponds to a single kinetic move that connects U_i and N_i. For a given pair of clusters U and N, many such micropathways usually exist. The sum over all these micropathways gives the intercluster transition rate:

$$k_{U\to N} = \sum_{U_i\to N_i} p_{U_i} k_{U_i\to N_i}; \quad k_{N\to U} = \sum_{N_i\to U_i} p_{N_i} k_{N_i\to U_i} \quad (4)$$

where p_{U_i} and p_{N_i} are the equilibrium fractional populations of U_i and N_i in the respective clusters. An important conclusion from eq 4 is that the intercluster transition rate $k_{U\rightleftharpoons N}$ and hence the overall folding rate is determined not only by the rate for each individual transition (micropathway) $k_{U_i\rightleftharpoons N_i}$ between two conformations but also by the fractional population of the "reactant" conformations (p_{U_i} and p_{N_i}) in the respective clusters.

Stochastic Method. Besides diagonalizing the rate matrix, another effective method to solve the master equation, especially for a large conformational ensemble, is the stochastic method.[28,29] In the stochastic method, the master equation is converted into a reaction probability density function: $P(\tau, \mu) = P_0(\tau)P_\mu r_\mu d\tau$, where $P(\tau, \mu)$ is the probability of the conformation transition μ occurring in the time interval $(t + \tau, t + \tau + d\tau)$, where $P_0(\tau)$ is the probability of no reaction occurring in the time interval $(t, t + \tau)$ and P_μ is the number of the different combinations of the "conformations" for conformation transition μ to occur. r_μ is the corresponding rate constant.[28,29] On the basis of the above reaction probability density function, Gillespie described a numerical algorithm to simulate the reaction process:

(1) Generate a series of independent random numbers n_u in the interval (0,1), $u = 1, 2, ..., M$. Here, M is the total number of conformation transitions.
(2) Calculate the transition time according to equation $\tau_\mu = (1/(P_\mu r_\mu)) \ln(1/n_\mu)$, $u = 1, 2, ..., M$. For example, for the transition from conformation i to conformation j, P_μ is equal to X_i, where X_i is the number of conformation i when the conformation transition μ occurs. We can calculate τ_μ straightforwardly from the equation.
(3) Choose the conformation transition μ with the smallest τ_μ as the actual transition at time t. The smallest τ_μ is the actual transition time τ. Set $t \leftarrow t + \tau$.
(4) Subtract the number of the "reactant" conformations by 1 and increase the number of the product by 1 for the selected conformation transition μ.
(5) Repeat steps 1−4 until all $P_\mu = 0$ or terminate the calculation at $t > t_{\text{stop}}$.

In Gillespie's algorithm, one needs to update and sort τ_μ in each iteration step, which is computationally expensive for a large number of reactions M. Gibson and Bruck described the next reaction method to avoid the step of updating and sorting τ_μ for all the conformation transitions.[30] The key points in the next reaction method[30] are (1) to change the relative time τ_μ to absolute time, (2) to store each reaction time τ_μ in a binary tree structure, and (3) only to update τ_μ for the conformation transitions in which P_μ is affected by the previous selected conformational transition. The next reaction method is significantly more efficient than Gillespie's algorithm. The details for implementing the next reaction method can be found in ref 30. In the study, we will use the next reaction method to simulate RNA folding kinetics.

III. Results and Discussion

Comparison with the Experiments. ***Folding Thermodynamics.*** Using real-time NMR spectroscopy, Wenter et al. measured the thermal stability and relaxation kinetic rates and the temperature dependence of the rate constants for two 20-nt hairpin-forming sequences GAAGGGCAACCUUCGGGUUG (sequence 1) and GACCGGAAGGUCCGCCUUCC (sequence

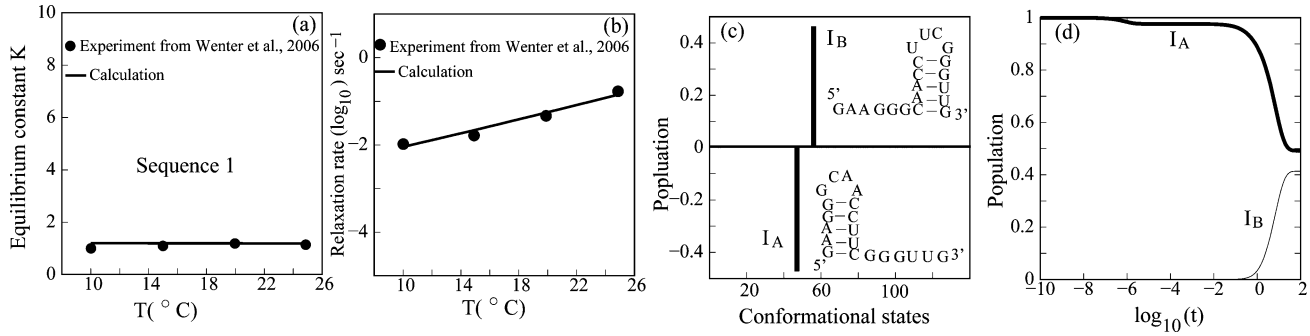

Figure 1. (a) Comparison of the equilibrium constants between the theoretical predictions (lines) and the experimental data (filled circles). (b) Theory−experiment comparisons for the relaxation rates. (c) The eigenvector for the slowest eigenvalue at $T = 25$ °C. (d) The population dynamics for the folding starting from conformation I_A at $T = 25$ °C. The experimental results are adopted from ref 22.

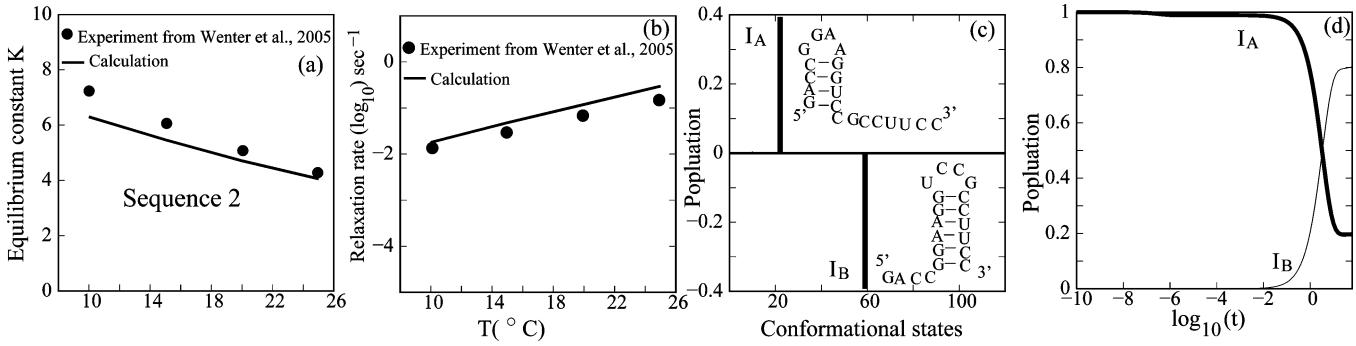

Figure 2. (a) The comparison of the equilibrium constants between the theoretical predictions (lines) and the experimental data (filled circles). (b) Theory−experiment comparison for the relaxation rates. The experimental results are adopted from ref 21. (c) The eigenvector for the slowest eigenvalue at $T = 25$ °C. (d) The population dynamics for the folding starting from conformation I_A at $T = 25$ °C. The experimental results are adopted from ref 21.

2).[21,22] Both sequences are found to fold into two alternative tetraloop hairpins I_A and I_B (see Figures 1c for sequence 1 and 2c for sequence 2). For sequence 1, the two structures have similar stabilities. For sequence 2, I_A, which contains a shorter stem, is significantly less stable than I_B.

Using the previously developed Vfold model,[31–33] we calculate the free energy difference $G_{I_A} - G_{I_B}$ between the two hairpin structures I_A and I_B in Figure 1c for sequence 1. Comparison with the experimental data[22] for the equilibrium constant $K = e^{-(G_{I_A}-G_{I_B})/k_BT}$ shows good agreement (Figure 1a). In the calculation, we use (10.2 kcal/mol, 28 cal/mol/K) as the (enthalpy, entropy) for the GCAA and UUCG tetraloops in I_A and I_B. The enthalpy and entropy parameters for the tetraloop formation are adopted from the previous thermodynamic measurement.[23]

For sequence 2, the hairpin structure I_A (Figure 2c) contains a GGAA tetraloop. The energetic parameter of the GGAA tetraloop is not available from the experiment. As an approximation, we fix the enthalpy of the GGAA tetraloop to 10.5 kcal/mol, which is close to the enthalpy of GCAA and UUCG tetraloops and fit the entropy parameter based on the experimentally measured equilibrium constant at $T = 25$ °C.[21] The fitted entropy parameter is 25 cal/mol/K. Using the fitted energetic parameters of the GGAA tetraloop, we calculate the equilibrium constants for other temperatures such as 10, 15, and 20 °C. As shown in Figure 2a, the comparison of the equilibrium constants with the experiments shows good agreement for a broad temperature range.[21]

Folding Kinetics. Using the master equation method, we can calculate the relaxation rate and the kinetic pathways. There are 135 conformations for each sequence. Diagonalizing the 135 × 135 rate matrix gives the eigenvectors and eigenvalues. Figure 3a shows the distribution of the eigenvalue spectrum at $T = 25$ °C. An apparent separation exists between λ_1 (the first nonzero eigenvalue) and λ_2 (the second nonzero eigenvalue). Thus, the relaxation kinetics is a single exponential process with the rate constant equal to λ_1. The predicted relaxation rate is in good agreement with the experiment (Figure 1b). The eigenvector for the lowest nonzero eigenvalue shows a two-state kinetic mode, corresponding exactly to the conversion between I_A and I_B populations (Figure 1c). The populational kinetics also shows a two-state transition between conformation I_A and I_B without any apparent intermediates (Figure 1d). The results agree with the experimental NMR observation.[22]

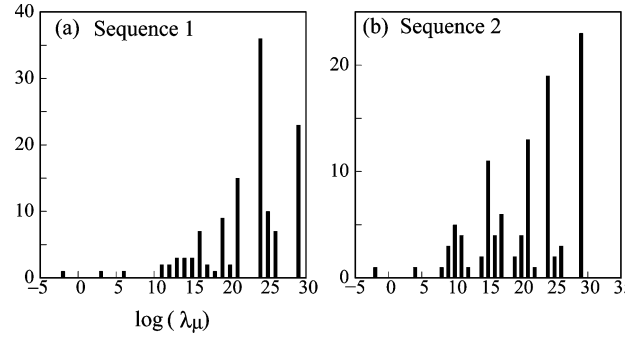

Figure 3. Histogram plots for the eigenvalue spectra at $T = 25$ °C for the two sequences. The *x*-axis denotes the conformation and the *y*-axis is the vector component for the corresponding conformation. The sequences in parts a and b are adopted from two experimental studies.[21,22]

For sequence 2, we predict similar kinetic behavior (Figures 2 and 3b): 2-state, single exponential kinetics for the transition between the two hairpin structures. The theoretical results are consistent with the experimental data.[21]

Comparison between Different Kinetic Models. ***Importance of Intraloop Base Pairs.*** In the previous kinetic cluster models, conformations with isolated base pairs were not treated

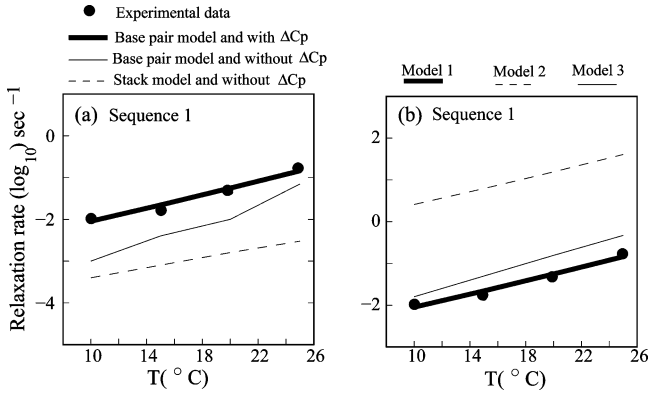

Figure 4. (a) Effects of ΔC_p and the single isolated base pair on the predicted relaxation rates. (b) Comparison of the relaxation rates between the three different rate models.

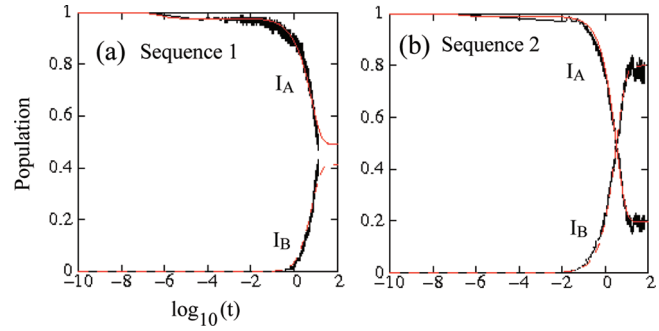

Figure 5. Populational dynamics starting from conformation I_A predicted from two different models: the master equation (dark) and the stochastic simulation (red). The sequences in parts a and b are adopted from two different experimental studies.[21,22] In the calculation, the temperature is set to 25 °C.

as a separate state because a lone base pair is not stable. Such an assumption is general valid for secondary structures.[6,19] However, for a tetraloop, a loop-closing base pair can enhance the formation of the noncanonical intraloop base pairs to stabilize the loop; thus, tetraloops closed by a base pair (prior to the formation of the loop-closing base stack) should be considered as a separate state in the conformational ensemble for kinetic modeling. Therefore, for the formation/disruption of tetraloops, we allow formation/disruption of a base pair in the kinetic move. The model leads to good agreement with the experiment (see Figure 4a). In contrast, as shown in Figure 4a, the base-stack-only conformational model, where only base stacks are explicitly considered in the kinetic moves, causes underestimation for the transition rate between the tetraloops I_A and I_B (in Figure 1b). This is due to the underestimation of the tetraloop stability.

ΔC_p Effect. Thermodynamics studies suggest that RNA folding may be accompanied with a change in the heat capacity ΔC_p,[34,35] resulting in a temperature dependence of the enthalpy and entropy parameters:

$$\Delta C_p = \left[\frac{\partial \Delta H}{\partial T}\right]_p = \left[T\frac{\partial \Delta S}{\partial T}\right]_p$$

Assuming a constant value for ΔC_p, we have

$$\Delta H(T) = \Delta H(37\ ^\circ\mathrm{C}) + \Delta C_p(T - 37),$$
$$\Delta S(T) = \Delta S(37\ ^\circ\mathrm{C}) + \Delta C_p \ln(T + 273)/310$$

where the parameters for ΔH(37 °C) and ΔS(37 °C) are the enthalpy and entropy parameters at 37 °C as given by the Turner rule.[37] As shown in Figure 4a, ΔC_p can have notable effects on the folding kinetics. Our predicted Arrhenius plot (Figure 4) based on $\Delta C_p = -40$ cal/mol/K/bp[34,36] gives good agreement with the experiment. The heat capacity change upon structural changes comes from not only the formation/disruption of the helices but also the formation/disruption of the tetraloop, which may involve desolvation/solvation of the loop nucleotides and the change in the solvent configuration.

Test with Different Rate Models. One of the key ingredients in the kinetic model is the evaluation of the rate constant for each elementary kinetic move such as the formation or disruption of a base stack. For the formation of a (stacked) base pair, we assume that, in the transition state, the participating nucleotides have positioned to form the base pair but have not yet "reacted" to form the stabilizing interactions; thus, the kinetic barrier is entropic (k_{close} in eq 3). For the disruption of a (stacked) base pair, we assume that, in the transition state, the stabilizing interactions have been disrupted while the base pairs have not yet been separated; thus, the kinetic barrier is enthalpic (k_{open} in eq 3).

Besides the above rate model ("model 1"), two other commonly used rate models exist, which relate the kinetic barrier to the free energy difference ΔG between the final and initial states:[13,14,38]

$$\begin{aligned}&\text{Model 2:}\quad k_{open} = k_0 e^{-\Delta G/k_B T},\quad k_{close} = k_0\\ &\text{Model 3:}\quad k_{open} = k_0 e^{-\Delta G/2k_B T},\quad k_{close} = k_0 e^{\Delta G/2k_B T}\end{aligned} \tag{5}$$

We use the three rate models to construct the respective rate matrices, from which we predict the kinetics based on the master equation. In models 2 and 3, we also consider the effect of heat capacity in the calculations of the temperature-dependent ΔG. As shown in Figure 4b, comparisons with experimental results indicate that model 1 gives better predictions than models 2 and 3 for the relaxation rate. Here, for models 2 and 3, the prefactor rate constant k_0, equal to 10^8 and 2.2×10^5 s^{-1} for models 2 and 3, respectively, are adopted from the previous studies.[13,38] The results show that models 2 and 3 overestimate the rate for the transition between states I_A and I_B. Moreover, the slope of the rate−temperature curve is slightly overestimated by model 3. The three models give similar overall folding rates except for a shift that can be potentially removed by rescaling of the prefactor k_0. Therefore, it might be possible that the model manifests their differences only on the k_0-independent folding properties such as the detailed pathways and kinetic intermediates.[8]

Master Equation Method and the Stochastic Model. In Figure 5a and b, we show the populational kinetics for states I_A and I_B predicted from the stochastic model (black lines). In the stochastic model, we use the next reaction method[30] to simulate the folding kinetics. The results from the stochastic model are nearly identical to the results from the exact master equation solution (red lines). The optimal theory−experiment fit gives scaling factor (rate constant) k_0 as $10^{-3.8}$ for the stochastic model. Nevertheless, we note that the stochastic model is much less efficient than the master equation solution: 5 days of CPU time vs 1 min of CPU time on a Dell EM64T cluster with the Intel (R) Xeon(R) 5150 (2.66 GHz) processor.

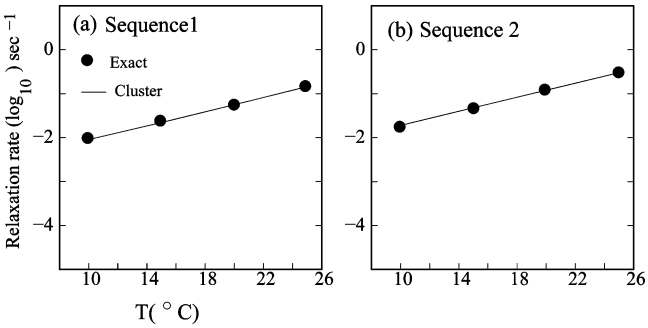

Figure 6. Comparison of the kinetic cluster method and the exact master equation method on the prediction of the relaxation rate. The sequences in parts a and b are adopted from two different experimental studies.[21,22]

Therefore, the stochastic model may not be feasible to study the long-time kinetic behavior.

Folding Pathways and Activation Energies. ***Folding Pathways.*** To identify the folding pathway using the kinetic cluster method, we first construct the clusters. The native base stacks that have large entropic cost $|\Delta S|$ to form are slow to fold. Similarly, the nonnative base stacks that have large enthalpic cost $|\Delta H|$ to form are slow to disrupt. These base stacks are potentially rate-limiting stacks. In addition, we identify the loop-closing base pairs that close GNRA, UNCG, and CUUG tetraloops. We use these pairs and the slow-forming/breaking native/nonnative base stacks to construct the kinetic clusters. Specifically, the different combinations of the presence and absence of these base pairs/stacks cause the different subsets of conformations and each subset forms a cluster. The transitions between two clusters correspond to the formation/disruption of a corresponding base pair or stack. As shown in eq 4, the conformational transitions (e.g., $U_i \rightarrow N_i$ in eq 4) that give the largest contributions ($p_{U_i} k_{U_i \rightarrow N_i}$ in eq 4) to the intercluster transition ($k_{U \rightarrow N}$ in eq 4) are identified as the dominant pathways.

For sequence 1 (Figure 1a), which has 135 conformations, we find five clusters: conformations with base stacks 4GG5−10CC11 (cluster C_1), 10CC11−16GG17 (cluster C_5), base pairs 5G−C10 (cluster C_2), 11C−G16 (cluster C_4), and without any of the above base stacks and base pairs (cluster C_3), respectively. Our master equation calculations show that the relaxation rate for the five clusters is exactly the same as that for the original complete conformational ensemble (see Figure 6a). The intercluster transitions show a single and reversible (dominant) pathway for $I_A \rightarrow I_B$ and $I_B \rightarrow I_A$ (see Figure 7a). From the intercluster rate constants, we find that the rate-limiting steps are the disruptions of base stacks 4GG5−10CC11 and 10CC11−16GG17 for $I_A \rightarrow I_B$ and $I_B \rightarrow I_A$, respectively.

To compute the rate constants for $I_A \rightarrow I_B$ and $I_B \rightarrow I_A$, we evaluate the rate for the corresponding dominant pathways using the following equation:[19,24,20]

$$k_n = k_1 r_{n-1} \prod_{p=1}^{n-2} \frac{r_p}{1 - r'_{p+1} r_p} \qquad (6)$$

where k_n is the overall folding rate for the pathway with n intermediates and the sum corresponds to the iterative rebounce processes and r_{n-1} and r'_n are the forward and backward (rebounce) probabilities for a molecule in states $n - 1$ and n, respectively:

$$r'_n = \frac{k_{n \rightarrow n-1}}{k_{n \rightarrow n+1} + k_{n \rightarrow n-1}}; \quad r_{n-1} = \frac{k_{n-1 \rightarrow n}}{k_{n-1 \rightarrow n} + k_{n-1 \rightarrow n-2}}$$

Equation 6 gives $k_{I_A \rightarrow I_B} \sim 0.065$ s^{-1} and $k_{I_B \rightarrow I_A} \sim 0.077$ s^{-1}, respectively. The results give the relaxation rate $k_{I_A \rightarrow I_B} + k_{I_B \rightarrow I_A} = 0.14$ s^{-1}, which agrees with the relaxation rate given by the lowest nonzero eigenvalue λ_1 of rate matrix.

For sequence 2 (Figure 2a), we classify the original 109 conformations in the complete ensemble into five clusters: conformations with base stacks 3CC4−9GG10 (cluster C_1), 9GG10−15CC16 (cluster C_4), base pairs 4C−G9 (cluster C_2), 10G−G15 (cluster C_5), and without any of the above base stacks and pairs (cluster C_3), respectively. The master equation for the five-cluster system gives exactly the same relaxation rate as the one for the original complete ensemble (see Figure 6b). Figure 7b shows the detailed folding pathways for $I_A \leftrightarrow I_B$. Equation 6 gives $k_{I_A \rightarrow I_B}$ and $k_{I_B \rightarrow I_A}$ equal to 0.26 and 0.06 s^{-1}, respectively. The relaxation rate is equal to $k_{I_A \rightarrow I_B} + k_{I_B \rightarrow I_A} = 0.32$ s^{-1} and is in good agreement with 0.30 s^{-1} from the exact master equation. The relaxation rate is mainly determined by $k_{I_A \rightarrow I_B}$.

The intercluster rate constants for the five-cluster system (C_1, ..., C_5 in Figure 7a) can provide further insights into the slow kinetics for the $I_A \leftrightarrow I_B$ conformational switch. From the rate constants shown in Figure 7a, the fractional population for the forward kinetics for clusters C_2 and C_3 are $r_2 = k_{C_2 \rightarrow C_3}/(k_{C_2 \rightarrow C_3} + k_{C_2 \rightarrow C_1}) = 0.11$ and $r_3 = k_{C_3 \rightarrow C_4}/(k_{C_3 \rightarrow C_4} + k_{C_3 \rightarrow C_2}) = 0.5$, respectively. Therefore, as a crude estimate, $k_{I_A \rightarrow I_B} \simeq k_{C_1 \rightarrow C_2} r_2 r_3 = 0.066$ s^{-1}, which is close to 0.065 s^{-1} from the exact master equation calculation (eq 6). The slow $I_A \rightarrow I_B$ transition is caused by the slow disruption of the loop-closing base stack 4GG5−10CC11 and the small forward folding probability for the single base pair-closed loop (I(0,1) in Figure 7a). Here, we label a structure as I(N, NN), where N is the number of "native" base pairs and NN is the number of "non-native" base pairs. We call a base pair "native" if it exists in structure I_B and "non-native" otherwise.

The kinetics for the conformational switch for sequence 2, as shown in Figure 7b, is similar to the one for sequence 1. The $I_A \rightarrow I_B$ transition is rate-limited by the disruption of the loop-closing base stack ($C_1 \rightarrow C_2$) and the small forward folding probability for C_2: $r_2 = k_{C_2 \rightarrow C_3}/(k_{C_2 \rightarrow C_3} + k_{C_2 \rightarrow C_1}) = 0.07$. The transition rate $k_{I_A \rightarrow I_B}$ can be estimated as $k_{C_1 \rightarrow C_2} \times r_2 = 0.25$ s^{-1}, which is close to 0.26 s^{-1} given by eq 6.

Activation Energies. From the temperature dependence of the relaxation rates, we calculate the activation energy (E_a) for the two-state kinetics: $E_a = -\mathrm{d} \ln(k)/\mathrm{d}(1/k_B T)$. The calculated activation energies for sequences 1 (Figure 1a) and 2 (Figure 2a) are 31 and 32 kcal/mol, respectively, which agree well with the experimental measurements (∼30 kcal/mol).[21,22] The activation energy is much smaller than the total enthalpy for the stable hairpin in the initial state. For instance, the total base-stacking enthalpy of hairpins I_A and I_B of Figure 1b is close to 49 kcal/mol based on Turner rules[37] and the energetic parameters for the tetraloops. The result suggests that the transition state is not the fully unfolded state. Wenter et al. proposed a pseudoknot transition state on the pathway.[21,22] From our kinetic studies, for the given sequences, we find that including or excluding pseudoknots does not cause notable differences in the predicted kinetics, although including the pseudoknots can affect the equilibrium population of states I_A and I_B for sequence 2. The current theoretical analysis suggests that pseudoknot structures probably have only a minor effect on the folding kinetics which makes dissection between associative and dissociative mechanisms difficult.

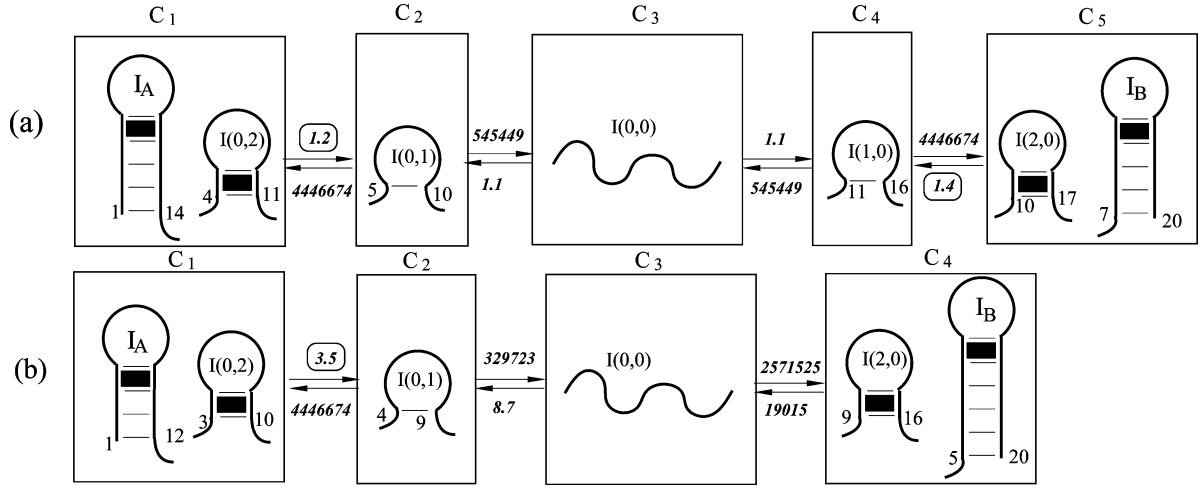

Figure 7. Predicted folding pathways for the transitions between conformations I_A and I_B at 25 °C. Sequences 1 (a) and 2 (b) are from two different experimental studies.[21,22] The intercluster transition rate (in s^{-1}) and the dominant micropathways between adjacent clusters are shown in the figure. Each structure in the figure is labeled as I(*N*,*NN*), in which *N* is the number of "native" base pairs and *NN* is the number of "non-native" base pairs. A base pair is "native" if it is found in structure I_B and "non-native" otherwise.

The folding/unfolding of the hairpin may be accompanied by heat capacity changes. Such a ΔC_p effect can influence the activation energy and cause the apparent theory−experiment difference in the activation energy. With ΔC_p, the rate for opening a base stack is (eq 3):

$$k = k_0 e^{(\Delta H_{stacks} + \Delta C_p(T-310))/k_B T}$$

where $\Delta H_{stack} < 0$ is the stacking enthalpy at $T = 37$ °C. As a result, the (apparent) activation energy becomes $E_a = -\Delta H_{stacks} + 310\Delta C_p$. A ΔC_p value of −0.04 kcal/mol/K would cause a decrease of 12.4 kcal/mol in the predicted activation energy. Therefore, the activation energy would be 49 kcal/mol − 12.4 kcal/mol = 36.6 kca/mol, which is close to the experimentally measured result.

IV. Conclusion

Using the master equation and kinetic cluster methods, we study the slow folding kinetics between two tetraloop hairpins. The predicted folding rates are in good agreement with the experiments. We find that the heat capacity change upon folding can have an important effect on the apparent activation energy. Such an effect may be more pronounced for tetraloops because the formation/disruption of a tetraloop involves noncanonical intraloop interactions and desolvation/solvation of the loop nucleotides. We note that the effect of heat capacity on RNA folding has been illustrated in previous studies.[39,40]

For the folding kinetics of secondary structures, which are stabilized by base stacking interactions, it may be appropriate to consider the formation/disruption of a base stack as a kinetic step. For the folding kinetics of a tetraloop, however, the formation/disruption of a loop-closing base pair (prior to the further stabilization by a loop-closing canonical base stack) should be considered as a kinetic move. This is because the loop closure is accompanied with intraloop interactions, causing a (kinetically metastable) state. This conclusion can be generalized to the folding kinetics of tertiary structures, where a base pair could juxtapose the different parts of the chain to cause the tertiary interactions and result in a kinetically metastable state.

For the conformational switches between the hairpins that we studied, the rate is limited by the disruption of the loop-closing base stack and the tetraloop (with the intraloop interactions). The detailed folding mechanism revealed in this study may be useful for understanding the slow folding kinetics of large RNAs, where folding and unfolding of tetraloop hairpins can be the fundamental steps in the overall structural changes. The current study is focused on the folding kinetics of small RNA structures in 1 M NaCl solution conditions. Further applications of the kinetic theory developed here would include the analysis and predictions for the folding kinetics of larger systems[41] with the influence of the different ions (Mg^{2+} and spermidine) in the solution.[42]

Acknowledgment. The research was supported by NSF through grant MCB0920411 (to S.-J.C.) and NIH grant GM063732 (to S.-J.C.). The Center of Biomolecular Magnetic Resonance is supported by the state of Hesse. H.S. is a member of the DFG-cluster of excellence: macromolecular complexes. The computations involved in this research were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).

References and Notes

(1) Chen, S.-J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646.
(2) Chen, S.-J. *Annu. Rev. Biophys.* **2008**, *37*, 197.
(3) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008**, *37*, 289.
(4) Porschke, D. *Biophys. Chem.* **1974**, *1*, 381.
(5) Ma, H.; Proctor, D. J.; Elzbieta, K.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523.
(6) Zhang, W. B.; Chen, S.-J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1931.
(7) Zhang, W. B.; Chen, S.-J. *Biophys. J.* **2006**, *90*, 765.
(8) Jung, J.; Van Orden, A. *J. Am. Chem. Soc.* **2006**, *128*, 1240.
(9) LeCuyer, K. A.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3373.
(10) Zarrinkar, P. P.; Williamson, J. R. *Science* **1994**, *265*, 918.
(11) Pan, J.; Woodson, S. A. *J. Mol. Biol.* **1998**, *280*, 597.
(12) Waldsich, C.; Pyle, A. M. *J. Mol. Biol.* **2008**, *375*, 572.
(13) Isambert, H.; Siggia, E. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6515.
(14) Flamm, C.; Fontana, W.; Hofacker, I. L.; Schuster, P. *RNA* **2000**, *6*, 325.
(15) Flamm, C.; Hofacker, I. L. *Monatsh. Chem.* **2008**, *139*, 447.

(16) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493.
(17) Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2008**, *130*, 9676.
(18) Cho, S. S.; Pincus, D. L.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17349.
(19) Cao, S.; Chen, S.-J. *J. Mol. Biol.* **2007**, *367*, 909.
(20) Zhao, P. N.; Zhang, W. B.; Chen, S.-J. *Biophys. J.* **2010**, *98*, 1617.
(21) Wenter, P.; Fürtig, B.; Hainard, A.; Schwalbe, H.; Pitsch, S. *Angew. Chem., Int. Ed.* **2005**, *44*, 2600.
(22) Wenter, P.; Fürtig, B.; Hainard, A.; Schwalbe, H.; Pitsch, S. *ChemBioChem* **2006**, *7*, 417.
(23) Proctor, D. J.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 14004.
(24) Zhang, W. B.; Chen, S.-J. *J. Chem. Phys.* **2003**, *119*, 8716.
(25) Cao, S.; Chen, S.-J. *Biophys. J.* **2009**, *96*, 4024.
(26) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13922.
(27) Konishi, Y.; Ooi, T.; Scheraga, H. A. *Biochemistry* **1982**, *21*, 4734.
(28) Gillespie, D. T. *J. Comput. Phys.* **1976**, *22*, 403.
(29) Gillespie, D. T. *J. Phys. Chem.* **1977**, *81*, 2340.
(30) Gibson, M. A.; Bruck, J. *J. Phys. Chem. A* **2000**, *104*, 1876.
(31) Cao, S.; Chen, S.-J. *RNA* **2005**, *11*, 1884.
(32) Cao, S.; Chen, S.-J. *Nucleic Acids Res.* **2006**, *34*, 2634.
(33) Cao, S.; Chen, S.-J. *RNA* **2009**, *15*, 696.
(34) Mikulecky, P. J.; Feig, A. L. *Biopolymers* **2006**, *82*, 38.
(35) Bourdélat-Parks, B. N.; Wartell, R. M. *Biochemistry* **2005**, *44*, 16710.
(36) Park, Y.-W.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1551.
(37) Serra, M. J.; Turner, D. H. *Methods Enzymol.* **1995**, *259*, 242.
(38) Liu, F.; Ou-Yang, Z.-C. *Biophys. J.* **2005**, *88*, 76.
(39) Draper, D. E.; Bukhman, Y. V.; Gluick, T. C. *Curr. Protoc. Nucleic Acid Chem.* DOI: 10.1002/0471142700.nc1103s02.
(40) Mikulecky, P. J.; Feig, A. L. *J. Am. Chem. Soc.* **2002**, *124*, 890.
(41) Fürtig, B.; Wenter, P.; Reymond, L.; Richter, C.; Pitsch, S.; Schwalbe, H. *J. Am. Chem. Soc.* **2008**, *129*, 16222.
(42) Fürtig, B.; Wenter, P.; Pitsch, S.; Schwalbe, H. *ACS Chem. Biol.* **2010**, *5*, 753.

JP107912S